A Note on Classification-Based Reasoning and Semi-Structured Objects

Amedeo Napoli

A note on classification-based reasoning and semi-structured objects

Amedeo Napoli
LORIA – UMR 7503
B.P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France
Email: napoli@loria.fr

Abstract

In this talk, we present a work in progress on the representation and manipulation of semi-structured data in an object-based representation environment. This research work is carried out in the field of knowledge representation and reasoning in order to build intelligent systems (according to artificial intelligence standards).

Object-based representation systems and classification-based reasoning are the representation and reasoning formalisms underlying this research work. An object-based representation system is based on a hierarchy $\mathtt{H} = (\mathtt{X}, \sqsubseteq)$ of classes, where classes in $\mathtt{X}$ are related by the subsumption relation $\sqsubseteq$ (a structural partial ordering). A class represents a real-world concept and is composed of a collection of attributes describing the characteristics and the behavior of the concept. Reasoning is based on inheritance (for knowledge sharing in $\mathtt{H}$) and classification, i.e. to insert a new class in $\mathtt{H}$, to determine the possible class of an object, to handle queries and to draw inferences.

Semi-structured data are data whose structure is not regular, heterogeneous, partial, has not a fixed format and quickly evolves. In our context, semi-structured data are represented as semi-structured objects, i.e. objects without classes, and they are manipulated by classification procedures. Semi-structured objects may be grouped within semi-structured classes that are composed of a conjunction and a disjunction of attributes. Semi-structured classes may be then classified within a hierarchy, on the basis of a subsumption relation $\sqsubseteq$, following an adaptation of the classical algorithm of classification-based reasoning (allowing to take into account disjunctions of attributes). Taking advantage of the hierarchical organization, semi-structured objects and classes may be then used for problem-solving purposes (drawing inferences on the data).

A parallel may be drawn between semi-structured classes and polythetic classes, introduced in the field of data analysis to take into account the irregularity of real-world concepts such as biological concepts. Semi-structured objects and classes may be exploited for several purposes, such as knowledge representation, knowledge discovery in data bases, text mining, integration of heterogeneous databases, and design of corporate memories.

Keywords: knowledge representation, reasoning formalisms, classification-based reasoning, problem solving, semi-structured data, polythetic classes, classification.